6 Pattern Mixture Models

A common theme underlying the methods we have discussed so far is that interest focuses on making inference on parameters in a parametric or semiparametric model for the ***full data*** Z given ***observed data*** $(R, Z_{(R)})$. As we discussed in Section 3.2, different approaches can be deduced by considering various ***factorizations*** of the density of the ***ideal full data*** (R, Z), which are of course unobservable.

A natural choice when a full data model $p_Z(z;\theta)$ has been posited is the ***selection model*** factorization

$$p_{R,Z}(r,z) = p_{R|Z}(r|z)p_Z(z). \tag{6.1}$$

An appealing feature of (6.1) is that the marginal distribution of the full data Z appears explicitly.

Accordingly, with posited models for the missingness mechanism and full data, we have

$$p_{R,Z}(r,z;\theta,\psi) = p_{R|Z}(r|z;\psi)p_Z(z;\theta), \tag{6.2}$$

where ψ is a nuisance parameter characterizing the missingness mechanism. In (6.2), if the parameters ψ and θ are assumed to be ***variation independent***, under MAR, as we have seen in Chapters 3-5, it is possible to estimate θ using likelihood, multiple imputation, or inverse probability weighted methods. Thus, under a selection model factorization and these assumptions, direct inference on the full data model of interest is possible. Of course, implementation of likelihood or inverse probability weighted methods can be challenging in some situations, and multiple imputation methods have the potential drawbacks we discussed in Chapter 4.

As we discussed in Section 3.2, another way of writing the joint density of the ideal full data is via the ***pattern mixture factorization***

$$p_{R,Z}(r,z) = p_{Z|R}(z|r)p_R(r). \tag{6.3}$$

In this chapter, we discuss inference based on adopting the factorization (6.3).

In contrast to the inferential methods based on the selection model factorization (6.1) that we have discussed to this point, methods based on (6.3) ***do not necessarily*** involve making the assumption of a MAR mechanism. Instead, they involve making other types of assumptions. Thus, as we will demonstrate shortly, they represent a ***fundamentally different*** approach to addressing inference in the presence of missing data. Not surprisingly, this approach has its own set of challenges.

6.1 Pattern mixture model framework

PATTERN MIXTURE MODEL: As we noted in Section 3.2, one may ***posit models*** for each of the components in the pattern mixture factorization (6.3). In (6.3), the first component $p_{Z|R}(z|r)$ is the density of the full data Z given the missing data pattern $R = r$. Usually, as noted in Section 3.2, R takes only a finite number of values r, and, in general, the distribution of Z for individuals having each pattern is expected to be different. Accordingly, for pattern r, consider the model

$$p_{R,Z}(r, z; \theta_r, \psi) = p_{Z|R}(z|r; \theta_r) p_R(r; \psi). \tag{6.4}$$

In (6.4), a separate model for $p_{Z|R}(z|r)$ is posited for each r, so we allow the parameter in the model to be r-dependent. Because R takes on only a finite number of values, $p_R(r; \psi)$ is simply the ***probability mass function*** for the ***discrete distribution*** of possible missingness patterns, represented by a single parameter ψ.

The parameter θ, comprising the distinct elements of θ_r for all observed missingness patterns r, and ψ in (6.4) have ***different interpretations*** from those in (6.2). In particular, θ in (6.4) is ***not*** the same as the parameter θ in (6.2) and is consequently ***not*** the parameter of direct interest. As we discuss below, if interest focuses on features of the distribution of the ***full data***, these must be deduced ***indirectly*** from (6.4).

The model (6.4) is referred to as a ***pattern mixture model*** because, for each ***pattern of missingness*** r, we model the conditional density of the full data Z given each pattern of missingness $R = r$. Because R is always observed, ψ can be estimated easily from the observed data.

Consequently, under the pattern mixture model (6.4), the main challenge is to estimate the parameters θ_r that characterize the model for the conditional density of Z given each missing data pattern.

DRAWBACK: One of the obvious drawbacks of pattern mixture models is that the marginal density of the full data is not ***explicitly represented*** in the formulation. Instead, the distribution of the full data is represented as the ***mixture***

$$p_Z(z; \theta, \psi) = \sum_r p_{Z|R}(z|r; \theta_r)\, p_R(r; \psi), \tag{6.5}$$

where, as usual, the sum in (6.5) is over all observed missingness patterns.

Accordingly, if features of the distribution of the full data are of inferential interest, these are represented only through a ***potentially complicated*** combination of θ and ψ.

ADVANTAGE: An advantage of adopting a pattern mixture model is that assumptions that allow ***identifiability*** are more explicit than with selection models, and estimation of the parameter θ in (6.4) is fairly straightforward, as we demonstrate shortly.

MONOTONE MISSINGNESS: Because pattern mixture models involve modeling the conditional distribution of the full data Z given the missingness pattern $R = r$ for ***each*** realized value r, such models are not useful if the number of missingness patterns is ***large***. This is because, in this case, the number of observations within a given pattern may be ***small***, so that the resulting estimates of these conditional distributions are not stable.

Consequently, pattern mixture model methods have not generally been applied to problems with ***arbitrary missingness***. Rather, the focus has been on their application in problems with ***monotone missingness***, where the number of patterns is limited. Accordingly, in this chapter, we consider the most familiar situation with monotone missingness, that of ***longitudinal data with dropout***.

As usual, write the full data as

$$Z = (Z_1, \ldots, Z_T),$$

where Z_j is observed at time t_j, $t_1 < \cdots < t_T$. Define $R = (R_1, \ldots, R_T)^T$ and the dropout indicator

$$D = 1 + \sum_{j=1}^{T} R_j,$$

so that $D = j$ corresponds to dropout at time t_j. Using notation defined in Chapter 1, with $D = j$, $Z_{(j)} = (Z_1, \ldots, Z_{j-1})$ is observed and $Z_{(\bar{j})} = (Z_j, \ldots, Z_T)$ is missing, $j = 2, \ldots, T+1$, where $D = T+1$ means $Z = (Z_1, \ldots, Z_T)$ is observed. The observed data are then $(D, Z_{(D)})$, and the sample data are $(D_i, Z_{(D_i)i})$, $i = 1, \ldots, N$. Assume that Z_1 is ***always observed***.

Under these conditions, for given j, we can write the pattern mixture factorization as

$$p_{Z,D}(z, j; \theta_j, \psi) = p_{Z|D}(z|j; \theta_j)\, p_D(j; \psi). \tag{6.6}$$

Analogous to (6.5), the ***marginal density*** of the full data Z is then the ***mixture***

$$p_Z(z; \theta, \psi) = \sum_{j=2}^{T+1} p_{Z|D}(z|j; \theta_j)\, p_D(j; \psi). \tag{6.7}$$

Contrast (6.6) with the selection model factorization, which instead involves a model for the ***dropout process*** $p_{D|Z}(j|z)$.

ASSUMPTIONS: Under the selection model factorization, we make assumptions directly on the dropout process. In Chapter 5, we made a ***MAR assumption***, which implies that $p_{D|Z}(j|z)$ satisfies

$$\text{pr}(D = j|Z) = \text{pr}(D = j|H_{j-1}) = \prod_{\ell=1}^{j-1}\{1 - \lambda_\ell(H_{\ell-1})\}\lambda_j(H_{j-1}), \tag{6.8}$$

$j = 2, \ldots, T+1$, depends only on the ***observed history*** $H_{j-1} = Z_{(j)} = (Z_1, \ldots, Z_{j-1})$ through time $j-1$.

We then ***modeled*** the ***dropout hazards*** $\lambda_j(H_{j-1})$ as $\lambda_j(H_{j-1};\psi_j)$, say, for each j and fit these models to the observed data. Letting ψ be the collection of ψ_j for $j = 2, \ldots, T+1$, from (6.8), these hazard models then induce models $p_{D|Z}(j|z;\psi)$ for each j depending on the components of ψ through ψ_j.

Instead, in the pattern mixture formulation (6.6), the ***marginal distribution*** of D is modeled ***directly***. Again, because there is a ***finite number*** of dropout patterns, this distribution is ***discrete***, so that $p_D(j;\psi)$ in (6.6) is a model for its ***probability mass function***. Thus, ψ in (6.6) can be estimated easily because D is always observed.

Of course, the ***catch*** is that we need to develop models $p_{Z|D}(z|j;\theta_j)$ for ***each*** different dropout pattern $D = j$. In doing so, we have to recognize that, as noted above, for any given pattern $D = j$, we observe only $Z_{(j)} = (Z_1, \ldots, Z_{j-1})$, whereas $Z_{(\bar{j})} = (Z_j, \ldots, Z_T)$ is missing.

To this end, note that, for $D = j$, we can write

$$p_{Z|D}(z|j) = p_{Z_{(j)}|D}(z_{(j)}|j)\, p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)}, j). \tag{6.9}$$

- The first term on the right hand side of (6.9), $p_{Z_{(j)}|D}(z_{(j)}|j)$, involves only ***observed data*** and thus is ***identified*** and can be ***modeled and fitted directly*** using the observed sample data.
- However, the second term on the right hand side of (6.9), $p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)}, j)$, involves the ***unobserved*** part of Z for $D = j$, and, accordingly, ***cannot*** be ***identified*** and thus modeled and fitted based on the observed sample data.

RESULT: The key challenge in the use of pattern mixture models is the need to make ***identifiability assumptions*** that allow us to deduce $p_{Z|D}(z|j)$, and in particular

$$p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)}, j), \qquad j = 1, \ldots, T$$

from the distribution of the observed data $(D, Z_{(D)})$.

We first examine an example where identifiability of the $p_{Z|D}(z|j)$ is achieved ***directly*** by making simple modeling assumptions that provide a way of ***extrapolating*** the observed data to unobserved data.

EXAMPLE: Longitudinal continuous outcome with dropout. Let the full data be

$$Z = (Y_1, \ldots, Y_T),$$

where Y_j are scalar, continuous outcomes observed at fixed time points t_j, $j = 1, \ldots, T$.

Suppose interest focuses on estimating the intercept and slope in an assumed ***simple linear regression model*** in time for the full data; i.e., we assume

$$E(Y_j) = \beta_0 + \beta_1 t_j, \tag{6.10}$$

and we are interested in inference on β_0 and β_1 in this (semiparametric) ***full data model***.

This would be straightforward with full data; however, suppose that some individuals ***drop out***. For simplicity, assume that ***all*** individuals have Y_1 and Y_2 observed (so the outcome is observed on all individuals at t_1 and t_2), but can drop out subsequently. Thus, we have possible ***dropout patterns*** $D = j = 3, \ldots, T+1$.

Suppose further that we are willing to assume that

$$E(Y_\ell|D = j) = \beta_{0j} + \beta_{1j} t_\ell, \quad \ell = 1, \ldots, T, j = 3, \ldots, T+1. \tag{6.11}$$

In (6.11), we assume that there is a linear trajectory ***within each dropout pattern***, but with intercept and slope that are ***pattern-dependent***. Note that (6.11) is a model for ***all*** Y_ℓ, $\ell = 1, \ldots, T$, even those that are ***missing*** under dropout pattern $D = j$. That is, (6.11) is a model for the ***full data*** Z given dropout pattern $D = j$ and depends on the ***pattern-specific parameter*** $\beta_j = (\beta_{0j}, \beta_{1j})$.

Under the foregoing assumptions,

$$\begin{aligned}
E(Y_j) &= E\{E(Y_j|D)\} \\
&= \sum_{j=3}^{T+1} E(Y_j|D = j)\text{pr}(D = j) \\
&= \sum_{j=3}^{T+1} (\beta_{0j} + \beta_{1j} t_j)\,\text{pr}(D = j) \qquad (6.12) \\
&= \left\{ \sum_{j=3}^{T+1} \beta_{0j}\,\text{pr}(D = j) \right\} + \left\{ \sum_{j=3}^{T+1} \beta_{1j}\,\text{pr}(D = j) \right\} t_j, \qquad (6.13)
\end{aligned}$$

where (6.12) follows from substituting the assumed model in (6.11).

Comparing (6.13) to (6.10), this formulation implies that the parameters of interest β_0 and β_1 in the ***full data model*** (6.10) are each a ***weighted average*** of the assumed pattern-specific intercepts and slopes; that is, we have the correspondences

$$\beta_0 = \sum_{j=3}^{T+1} \beta_{0j}\,\text{pr}(D=j), \qquad \beta_1 = \sum_{j=3}^{T+1} \beta_{1j}\,\text{pr}(D=j).$$

Each of the pattern-specific intercepts and slopes can be estimated easily from the observed data within each pattern; that is, for each j, β_{0j} and β_{1j} can be estimated by fitting the model (6.11) using the observed data on individuals for whom $D = j$. Because all individuals have Y_1 and Y_2 observed, this is possible for all dropout patterns. Denote these ***pattern-specific*** estimates by $(\widehat{\beta}_{0j}, \widehat{\beta}_{1j})$, $j = 3, \ldots, T+1$.

We can also estimate $\text{pr}(D = j)$ for $j = 3, \ldots, T+1$ by the ***observed proportions*** of individuals $\widehat{\psi}_j$, say, exhibiting each dropout pattern.

Then from (6.13), we can estimate β_0 and β_1 by

$$\widehat{\beta}_0 = \sum_{j=3}^{T+1} \widehat{\beta}_{0j}\,\widehat{\psi}_j, \qquad \widehat{\beta}_1 = \sum_{j=3}^{T+1} \widehat{\beta}_{1j}\,\widehat{\psi}_j.$$

NONIDENTIFIABLE ASSUMPTION: This approach to estimation has considerable appeal, as the resulting estimators for the parameters of interest in the full data model are ***easy to implement***. ***However***, it is important to recognize that this approach is predicated on a ***critical assumption***, namely, that in (6.11),

$$E(Y_\ell | D = j) = \beta_{0j} + \beta_{1j} t_\ell, \quad \ell = 1, \ldots, T, j = 3, \ldots, T+1.$$

- The assumption of a ***linear relationship*** within a dropout pattern $D = j$ implies the belief that the observed data $Y_1, \ldots, Y_{j-1}$ under that pattern can be ***extrapolated*** to the data $Y_j, \ldots, Y_T$ that are ***not observed***.
- Clearly, this an assumption that is ***not identifiable*** from the observed data.

RESULT: This example illustrates in a simple setting the type of unverifiable assumption that might be made in a ***pattern mixture model*** framework to ***identify*** the densities $p_{Z|D}(z|j)$. Here, the structure of the problem is such that parameters in the full data model of interest can be deduced directly from the resulting ***mixture*** (6.7).

In more general settings, such a simple approach ***may not*** be possible. In the next section, we discuss more general strategies that involve making ***identifying assumptions*** for the unobservable distributions

$$p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)},j) \quad j = 1,\ldots,T \tag{6.14}$$

from the distribution of the observed data.

RELATIONSHIP TO MISSINGNESS MECHANISM: The identifying assumption that is made, along with the observable distributions

$$p_{Z_{(j)}|D}(z_{(j)}|j) \quad \text{and } p_D(j),$$

induce a probability distribution for the ***missingness*** (***dropout***) ***mechanism***,

$$p_{D|Z}(j|z),$$

in some possibly ***complicated fashion***.

That is, using (6.9) and Bayes rule, the dropout mechanism can be represented as

$$p_{D|Z}(j|z) = \frac{p_{Z_{(j)}|D}(z_{(j)}|j)\, p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)},j)p_D(j)}{\sum_{j=2}^{T+1} p_{Z_{(j)}|D}(z_{(j)}|D=j)\, p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)},D=j)\,\text{pr}(D=j)}. \tag{6.15}$$

Consequently, within the pattern mixture model framework, it is evident that assumptions made to identify the unobservable distributions in (6.14) ***will not necessarily*** lead to the MAR assumption. This is in contrast to working within a ***selection model*** framework, where making the MAR assumption ***explicitly*** is the natural way to achieve identifiability.

Of course, neither the assumptions made on the unobservable distribution $p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)},j)$ in a pattern mixture model approach nor the MAR assumption made in a selection model approach can be ***validated*** from the observed data.

6.2 Modeling strategies for pattern mixture models

We now consider general strategies for developing models $p_{Z|D}(z|j;\theta_j)$, say, for the density $p_{Z|D}(z|j)$ for each j based on various types of ***identifying assumptions***.

Given models $p_{Z|D}(z|j;\theta_j)$ so constructed, a ***mixture model*** for the ***full data*** is then induced as in (6.7) as

$$p_Z(z;\theta,\psi) = \sum_{j=2}^{T+1} p_{Z|D}(z|j;\theta_j)\, p_D(j;\psi). \tag{6.16}$$

In (6.16), $p_D(j;\psi)$ can be represented by the ***probability mass function***

$$\text{pr}(D = j) = \psi_j, \quad j = 2, \ldots, T+1, \quad \sum_{j=2}^{T+1} \psi_j = 1, \tag{6.17}$$

and the ψ_j estimated by the sample proportions of individuals with $D = j$.

EXAMPLE: Throughout this section, we illustrate in the context of a simple example. Suppose that the ***full data*** are

$$Z = (Y_1, Y_2, Y_3)$$

so that $T = 3$ and $T + 1 = 4$, where Y_j are scalar, continuous outcomes observed at fixed time points t_j, $j = 1, 2, 3$. Let $Y = (Y_1, Y_2, Y_3)^T$. Interest focuses on estimating the mean $E(Y) = \mu = (\mu_1, \mu_2, \mu_3)^T$ and covariance matrix $\text{var}(Y) = \Sigma$.

Here, the models $p_{Z|D}(z|j;\theta_j)$, $j = 2, 3, 4$, are of the form

$$p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|2;\theta_2), \quad p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|3;\theta_3), \quad p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|4;\theta_4). \tag{6.18}$$

Substituting (6.18) and (6.17) in (6.16), the implied ***full data model*** is

$$p_{Y_1,Y_2,Y_3}(y_1, y_2, y_3;\theta,\psi) = \sum_{j=2}^{4} p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|j;\theta_j)\, \psi_j. \tag{6.19}$$

The parameters of interest, μ and Σ, would need to be ***derived*** from (6.19) and would clearly be possibly ***complicated functions*** of θ and ψ.

For example, letting

$$\nu_j(\theta_j) = \int (y_1, y_2, y_3)^T\, p_{Y_1,Y_2,Y_3}(y_1, y_2, y_3|j;\theta_j)\, dy_1\, dy_2\, dy_3,$$

be the mean vector for the jth component in the mixture (6.19), it follows from (6.19) that

$$\mu = \sum_{j=2}^{4} \nu_j(\theta_j)\psi_j.$$

Calculation of an expression for Σ is more involved (try it).

MODELING $p_{Z|D}(z|j)$ ***UNDER IDENTIFYING ASSUMPTIONS:*** From (6.9), we have for each $j = 2, \ldots, T+1$

$$p_{Z|D}(z|j) = p_{Z_{(j)}|D}(z_{(j)}|j)\, p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)}, j). \tag{6.20}$$

(i) The first term on the right hand side of (6.20) depends on observed data, so can be ***modeled directly*** based on the observed sample data; e.g., by a parametric model

$$p_{Z_{(j)}|D}(z_{(j)}|j;\xi_j), \quad j = 2, \ldots, T+1, \tag{6.21}$$

say. For each j, this model can be posited and fitted using the observed sample data on $Z_{(j)}$ from individuals for whom $D = j$; i.e., who dropped out at t_j and thus have $Z_{(j)} = (Z_1, \ldots, Z_{j-1})$ observed.

(ii) The second term on the right hand side of (6.20) is ***not identifiable*** from the observed data without some type of ***identifiability assumption***.

EXAMPLE, (i): The required models in (6.21) are

$$p_{Y_1|D}(y_1|2;\xi_2), \quad p_{Y_1,Y_2|D}(y_1,y_2|3;\xi_3), \quad p_{Y_1,Y_2,Y_3|D}(y_1,y_2,y_3|4;\xi_4), \tag{6.22}$$

Each parametric model in (6.22) can be developed and fit using the observed data from individuals exhibiting each ***dropout pattern*** $j = 2, 3, 4$.

Proposed modeling strategies for (ii) are based on writing the second term in (6.20) as

$$p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)}, j) = \prod_{\ell=j}^{T} p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)}, j). \tag{6.23}$$

From (6.23), the challenge then boils down to making assumptions on the ***unobservable*** conditional densities

$$p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)}, j), \quad \ell = j, \ldots, T, \;\; j = 2, \ldots, T. \tag{6.24}$$

We describe three main types of assumptions on (6.24). Section 3.6 and Chapter 16 of Molenberghs and Kenward (2007) present additional details and accounts of implementation.

CCMV: Complete case missing values. This approach is proposed by Little (1993). Here, it is assumed that (6.24) can be written as

$$p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},j) = p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},T+1), \quad \ell = j,\ldots,T, \quad j = 2,\ldots,T. \tag{6.25}$$

That is, in (6.25), it is assumed that the densities for the missing part of Z given the observed for each pattern $D = j$, $j = 2,\ldots,T$, are the ***same*** as those for the ***complete cases*** for whom $D = T+1$. Note that the densities $p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},T+1)$ are ***always observable***, so can be modeled and fitted based on the observed sample data.

Under the CCMV approach, then, one must deduce models for

$$p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},T+1), \quad \ell = 2,\ldots,T. \tag{6.26}$$

One possibility would be to posit ***directly*** parametric models $p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},T+1;\vartheta_\ell)$, say, for $\ell = 2,\ldots,T$. However, note that, given that we have ***already specified*** identifiable models for $p_{Z_{(j)}|D}(z_{(j)}|j;\xi_j)$ as in (6.21), the models for (6.26) should be ***compatible*** with these identifiable models.

We demonstrate how this is accomplished in the example.

EXAMPLE, (ii), CCMV: From (6.26), we require models for

$$p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,4), \qquad p_{Y_2|Y_1,D}(y_2|y_1,4). \tag{6.27}$$

From (6.22), we have already posited a model

$$p_{Y_1,Y_2,Y_3|D}(y_1,y_2,y_3|4;\xi_4). \tag{6.28}$$

In principle, we can thus obtain

$$\begin{aligned} p_{Y_1,Y_2|D}(y_1,y_2|4;\xi_4) &= \int p_{Y_1,Y_2,Y_3}(y_1,y_2,y_3|4;\xi_4)\,dy_3, \\ p_{Y_1|D}(y_1|4;\xi_4) &= \int p_{Y_1,Y_2|D}(y_1,y_2|4;\xi_4)\,dy_2, \end{aligned} \tag{6.29}$$

from whence it follows that the required models in (6.27) can be deduced as

$$\begin{aligned} p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,4;\xi_4) &= \frac{p_{Y_1,Y_2,Y_3|D}(y_1,y_2,y_3|4;\xi_4)}{p_{Y_1,Y_2|D}(y_1,y_2|4;\xi_4)}, \\ p_{Y_2|Y_1,D}(y_2|y_1,4;\xi_4) &= \frac{p_{Y_1,Y_2|D}(y_1,y_2|4;\xi_4)}{p_{Y_1|D}(y_1|4;\xi_4)}. \end{aligned} \tag{6.30}$$

Thus, under the CCMV assumption given a model (6.28), models for the conditional densities in (6.27) are ***automatically induced*** from the model (6.28).

From (6.20) and (6.23), models for $p_{Z|D}(z|j)$, $j = 2, 3, 4$, are thus given by

$$p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|2; \xi_2, \xi_4) = p_{Y_1|D}(y_1|2; \xi_2) \left\{ p_{Y_2|Y_1,D}(y_2|y_1, 4; \xi_4)\, p_{Y_3|Y_2,Y_1,D}(y_3|y_2, y_1, 4; \xi_4) \right\}, \tag{6.31}$$

$$p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|3; \xi_3, \xi_4) = p_{Y_1,Y_2|D}(y_1, y_2|3; \xi_3) p_{Y_3|Y_2,Y_1,D}(y_3|y_2, y_1, 4; \xi_4). \tag{6.32}$$

and $p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|4; \xi_4)$ in (6.22). The expressions on the right hand sides of (6.31) and (6.32) as well as this last density are all in terms of the models in (6.22) based on the observed data. Note that we thus have

$$\theta_2 = (\xi_2^T, \xi_4^T)^T, \quad \theta_3 = (\xi_3^T, \xi_4^T)^T, \quad \theta_4 = \xi_4,$$

and we can define $\theta = (\xi_2^T, \xi_3^T, \xi_4^T)^T$ to be the vector of distinct elements of $\theta_2, \theta_3, \theta_4$.

Substituting all of the above in (6.19) yields resulting ***full data mixture model***.

MULTIVARIATE NORMALITY: As a specific example, one could assume that each of the densities in (6.22), namely,

$$p_{Y_1|D}(y_1|2; \xi_2), \quad p_{Y_1,Y_2|D}(y_1, y_2|3; \xi_3), \quad p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|4; \xi_4),$$

is ***normal***. In this case, ξ_j would comprise mean and distinct variance and covariance parameters; i.e., in obvious notation, $\xi_2 = (\mu^{(2)}, \sigma^{(2)2})^T$, ξ_3 comprises $\mu^{(3)} = (\mu_1^{(3)}, \mu_2^{(3)})^T$ and the distinct elements of a (2×2) covariance matrix $\Sigma^{(3)}$, and ξ_4 comprises $\mu^{(4)} = (\mu_1^{(4)}, \mu_2^{(4)}, \mu_3^{(4)})^T$ and the distinct elements of a (3×3) covariance matrix $\Sigma^{(4)}$.

It then follows by properties of the multivariate normal distribution, including the results for ***conditional densities*** given in (4.35) and (4.36) in Chapter 4, that all the components on the right hand sides of (6.31) and (6.32) are normal densities that can be expressed in terms of ξ_2, ξ_3, ξ_4. It further follows that each of

$$p_{Y_1,Y_2,Y_3}(y_1, y_2, y_3|j; \theta_j).$$

where θ_j are defined as above, are ***normal densities***.

Thus, adopting normality assumptions results in a ***full data model*** that is in the form of a ***mixture*** of normal densities.

REMARK: Ordinarily, in practice, a data analyst ***would not***, indeed ***could not***, specify a normal mixture depending on ***observed dropout patterns*** as a model for the ***full data*** *a priori*; i.e., prior to seeing the observed data.

NCMV: Neighboring case missing values. In this approach, it is assumed instead that

$$p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},j) = p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},\ell+1), \quad \ell = j,\ldots,T, \;\; j = 2,\ldots,T. \tag{6.33}$$

Here, the assumption is that the densities for each dropout pattern $D = j$, $j = 2,\ldots,T$ are the same as those corresponding to the first available dropout pattern d for which $p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},d)$ is observable; namely, $D = \ell + 1$.

As with CCMV, rather than posit models ***directly*** for $p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},\ell+1)$ for each j and $\ell = j,\ldots,T$, models ***compatible*** with those in (6.21) can be deduced.

We demonstrate in the example.

EXAMPLE, (ii), NCMV: From (6.33), it is assumed that

$$p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,2) = p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,3) = p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,4),$$

$$p_{Y_2|Y_1,D}(y_2|y_1,2) = p_{Y_2|Y_1,D}(y_2|y_1,3).$$

We thus require models for

$$p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,4), \quad p_{Y_2|Y_1,D}(y_2|y_1,3). \tag{6.34}$$

As for CCMV, from (6.22), we have already posited a model $p_{Y_1,Y_2,Y_3|D}(y_1,y_2,y_3|4;\xi_4)$; thus, by the same calculations as in (6.29) and (6.30), a model

$$p_{Y_3|Y_2,Y_1,D}(y_3|y_2,y_1,4;\xi_4)$$

is induced.

To deduce a model for the second term in (6.34), note that from (6.22), we have already posited a model

$$p_{Y_1,Y_2,D}(y_1,y_2|3;\xi_3).$$

By entirely similar calculations, we thus have

$$\begin{aligned} p_{Y_1|D}(y_1|3;\xi_3) &= \int p_{Y_1,Y_2}(y_1,y_2|3;\xi_3)\,dy_2, \\ p_{Y_2|Y_1,D}(y_2|y_1,3;\xi_3) &= \frac{p_{Y_1,Y_2|D}(y_1,y_2|3;\xi_3)}{p_{Y_1|D}(y_1|3;\xi_3)}. \end{aligned}$$

Thus, as with CCMV, the required models in (6.34) are ***automatically induced*** by those in (6.22).

Again, from (6.20) and (6.23), models for $p_{Z|D}(z|j)$, $j = 2, 3, 4$, are thus given by

$$\begin{aligned} p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|2; \xi_2, \xi_3, \xi_4) &= p_{Y_1|D}(y_1|2; \xi_2) \left\{ p_{Y_2|Y_1,D}(y_2|y_1, 3; \xi_3)\, p_{Y_3|Y_2,Y_1,D}(y_3|y_2, y_1, 4; \xi_4) \right\}, \\ p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|3; \xi_3, \xi_4) &= p_{Y_1,Y_2|D}(y_1, y_2|3; \xi_3) p_{Y_3|Y_2,Y_1,D}(y_3|y_2, y_1, 4; \xi_4). \end{aligned}$$

and $p_{Y_1,Y_2,Y_3|D}(y_1, y_2, y_3|4; \xi_4)$. Thus,

$$\theta_2 = (\xi_2^T, \xi_3^T, \xi_4^T)^T, \quad \theta_3 = (\xi_3^T, \xi_4^T)^T, \quad \theta_4 = \xi_4,$$

and $\theta = (\xi_2^T, \xi_3^T, \xi_4^T)^T$.

As with CCMV, substituting the above expressions in (6.19) yields resulting ***full data mixture model***. Moreover, assuming that the densities in (6.22) are ***normal*** would lead to a similar ***normal mixture model*** for the full data.

REMARK: Under each of the CCMV and NCMV assumptions, it is possible, by substituting in (6.15), to derive the ***dropout mechanism*** $p_{D|Z}(j, z)$ ***induced*** by these assumptions. Via such calculations, it can be shown (try it) that neither CCMV nor NCMV results in an induced mechanism that is MAR in general.

GENERAL FORMULATION: The CCMV and NCMV approaches to identifiability are ***special cases*** of the following ***general formulation***.

Write the required ***nonidentifable*** conditional densities in (6.24), namely

$$p_{Z_\ell|Z_{(\ell)},D}(z_{(\ell)}, j), \quad \ell = j, \ldots, T, \quad j = 2, \ldots, T,$$

or, equivalently, $\ell = 2, \ldots, T$, $j = 1, \ldots, \ell - 1$, as a ***weighted average*** of the corresponding conditional densities across all dropout patterns where that conditional density is ***identifiable***; i.e.,

$$p_{Z_\ell|Z_{(\ell)},D}(z_{(\ell)}, j) = \sum_{d=\ell+1}^{T+1} \omega_{\ell d}\, p_{Z_\ell|Z_{(\ell)},D}(z_{(\ell)}, d), \qquad \sum_{d=\ell+1}^{T+1} \omega_{\ell d} = 1. \tag{6.35}$$

- CCMV is the special case of (6.35) with

$$\omega_{\ell d} = 0, \quad d = 1, \ldots, T; \qquad \omega_{\ell, T+1} = 1.$$

- NCMV is the special case of (6.35) with

$$\omega_{\ell,\ell+1} = 1, \qquad \omega_{\ell d} = 0, \quad d \neq \ell + 1.$$

ACMV: Available case missing values. Another special case of (6.35) proposed by Molenberghs et al. (1998) and referred to by these authors as ***available case missing values*** (ACMV), involves a particular set of non-zero weights $\omega_{\ell d}$. Recalling that $\psi_j = \text{pr}(D = j)$,

$$\omega_{\ell d} = \frac{\psi_d \, p_{Z_\ell | Z_{(\ell)}, D}(z_{(\ell)}, d)}{\sum_{p=\ell+1}^{T+1} \psi_p \, p_{Z_\ell | Z_{(\ell)}, D}(z_{(\ell)}, p)}, \quad d = \ell + 1, \ldots, T + 1. \tag{6.36}$$

RELATIONSHIP TO MAR: It turns out, as shown by Molenberghs et al. (1998), that taking a pattern mixture perspective as we have here and achieving identifiability by using the weights (6.36) in (6.35) is ***equivalent*** to making the assumption of MAR; see this article for a detailed proof. Indeed, establishing equivalence between adopting a ***pattern mixture model*** and ***identifiability assumptions*** and ***MAR*** was the motivation for the form of the weights in (6.36).

SUMMARY: The foregoing developments demonstrate that implementation of a pattern mixture model in practice involves development and fitting of a series of models, namely, those in (6.21) that ***are identifiable*** from the observed data ***and*** those in (6.24) dictated by the identifiability assumption adopted. Accordingly, implementation can be intensive in practical settings that are more complex than our simple illustrative example.

6.3 Implementing pattern mixture models

As discussed in Section 16.4 of Molenberghs and Kenward (2007), in general, if one takes a pattern mixture approach, adopting ***identifying assumptions*** such as CCMV, NCMV, or ACMV as discussed in the previous section, estimators for full data quantities of interest are usually ***not readily available*** but instead must be deduced from the induced ***full data mixture model***. Moreover, there is no obvious way to derive ***measures of uncertainty*** (e.g., standard errors and confidence intervals) for these quantities.

An approach that has been advocated in the literature to address these issues is to ***combine*** the use of pattern mixture models with ***multiple imputation***. In Section 16.5 of Molenberghs and Kenward (2007), this general strategy is presented. A summary of the key steps is as follows.

1. Posit and fit models

$$p_{Z_{(j)}|D}(z_{(j)}|j;\xi_j), \quad j = 2, \ldots, T+1,$$

in (6.21). These models presumably would be chosen so that estimators for the parameters ξ_j and approximate measures of uncertainty for them (e.g., standard errors, asymptotic covariance matrices) are ***readily available***.

2. Select an ***identification approach***, e.g., CCMV, NCMV, or ACMV.

3. Using this identification approach, determine the conditional densities of the ***unobserved outcomes given the observed***, that is

$$p_{Z_{(\bar{j})}|Z_{(j)},D}(z_{(\bar{j})}|z_{(j)},j) = \prod_{\ell=j}^{T} p_{Z_\ell|Z_{(\ell)},D}(z_\ell|z_{(\ell)},j), \quad j = 2, \ldots, T. \tag{6.37}$$

4. Using ***multiple imputation***, based on this formulation, draw multiple imputations for the unobserved components given the observed and the pattern-specific densities (6.37). It is necessary to use ***proper*** imputation methods; that is, it is also required to make draws from the ***predictive distribution*** of the ξ_j as well as from the conditional densities on the right hand side of (6.37).

5. Analyze the multiply imputed data sets using a ***model and method of choice***. For example, one could adopt a ***selection model*** or another pattern mixture model or any other model of interest.

6. Inference is then carried out using the standard multiple imputation approach, with standard errors obtained via ***Rubin's formula***.

See Chapter 16 of Molenberghs and Kenward (2007) for a detailed account of use of this strategy in several applications.

6.4 Discussion

As we have demonstrated in this chapter, inference based on pattern mixture models does not involve directly making the assumption of a MAR mechanism, as is the case for methods that are based on a selection model framework. Instead, the pattern mixture approach involves making other types of (***untestable***) ***identifiability*** assumptions. CCMV and NCMV are two relatively straightforward approaches to doing this. The ACMV formulation demonstrates that there is an ***intersection*** of these

two very different perspectives on modeling. However, except in this case, in general, the dropout mechanism that is implied by a pattern mixture model is ***not transparent***.

A prominent feature of pattern mixture models is that quantities associated with the ***full data distribution***, which would be ***modeled explicitly*** in a selection model approach, are not readily available and must be ***derived*** from the implied full data model. As shown in (6.16) this model is a ***mixture*** of models for the full data ***given*** the observed missing data patterns. In general, this induced model ***would not*** correspond to a full data model that a data analyst might posit ***explicitly***, as the induced model depends on the probability mass function of possible missing data patterns, which would not be known *a priori*.

Molenberghs and Kenward (2007) note that pattern mixture models can yield useful insights in practical applications and thus can be a valuable complement to the selection-based approaches discussed in Chapters 3-5. The same can be said in the context of ***sensitivity analysis***, discussed in the next chapter.